ELSEVIER

Gene 215 (1998) 471–476

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework

P. Jirholt [a], M. Ohlin [a], C.A.K. Borrebaeck [a], E. Söderlind [a,b,*]

[a] *Department of Immunotechnology, Lund University, Lund, Sweden*
[b] *BioInvent Therapeutic AB, Lund, Sweden*

Received 31 March 1998; received in revised form 2 June 1998; accepted 2 June 1998; Received by T. Gojobori

Abstract

A novel approach in molecular design is presented, where in vivo formed complementarity determining regions (CDR) from antibody genes were shuffled into a specific framework region. A synthetic gene library of soluble VH-fragments was created and the complexity of the library was determined by sequencing. The synthetic genes were diverse and contained random combinations of CDR from different germlines. All CDR were randomised in one step and by using in vivo formed CDR, the length, sequence and combination were varied simultaneously. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Immunoglobulin germline gene; Synthetic gene library; Primer design; Gene assembly; CDR shuffling

1. Introduction

Combinatorial biology provides an efficient way of creating large molecular libraries, and it has been applied, in particular, to the V-region of antibody genes. Since the probability of finding antibodies with a certain specificity is higher in an antibody library which contains a large number of individual clones, generation of diversity is a key element. The target segments for introduction of diversity are the complementarity determining regions (CDR) of antibody genes and different ways have been used, such as PCR amplification of V-regions using randomised primers covering the CDR (Hoogenboom and Winter, 1992; Barbas III et al., 1992; Griffiths et al., 1994) or the use of V-regions from in vivo immunised donors. Error-prone PCR or bacterial mutator strains have also been used to introduce mutation in a more unrestricted way (for review see Hayden et al., 1997). It is also possible to synthetically construct the entire antibody V-regions in vitro, using overlapping oligonucleotides subsequently assembled into full-length genes. This concept was demonstrated recently with the assembly of antibody light-chain gene libraries, using totally randomised CDR (Hayashi et al., 1994) or by retaining the canonical residues of each CDR while randomising the remaining parts of these sequences during the in vitro synthesis (Söderlind et al., 1995), which will minimise adverse structural effects. The approach using synthetic V-regions assembled in vitro has been demonstrated to yield a fully functional anti-FITC single-chain antibody fragment based on the DP-47 and DPL-3 germline gene sequences (Kobayashi et al., 1997), thus paving the way for construction of more elaborate synthetic antibody libraries. Similarly, the concept of molecular evolution through combination of DNA segments between related genes has been demonstrated by other laboratories (Crameri and Stemmer, 1995; Stemmer, 1994; Zhao et al., 1998).

In the present study we have developed the concept of synthetic antibody design to allow the introduction of greater variability into a library, based on the utilisation of gene diversity created in vivo. This was achieved by shuffling CDR isolated in vivo into a single preselected human framework (master framework), and we used a single-domain V_H antibody fragment to demonstrate this concept. A molecular library of single-domain V_H antibody fragments was constructed on the DP-47 master framework (FR) that has been camelised in three

* Corresponding author. Present address: Department of Immunotechnology, Lund University, P.O. Box 7031, S-220 07 Lund, Sweden. Tel: +46-2224322; Fax: +46-2224200; e-mail: eskil.soderlind@immun.lth.se

Abbreviations: CDR, Complementarity determining region(s); FR, framework region(s); PCR, polymerase chain reaction; V, variable; V_H, variable region of the immunoglobulin heavy chain.

PII: S0378-1119(98)00317-5

positions (Davies and Riechmann, 1995) and using CDR1–3 originating from different V genes formed in vivo. This approach of CDR shuffling allowed, for the first time, a combination of CDR originating from different antibody genes into one antibody fragment gene, while also sampling the entire pool of CDR. An enormous diversity can be introduced into V-regions by this approach which, consequently, has the potential for creating new binding specificities as well as for evolving antibody affinities in vitro in fragments with known specificities.

2. Materials and methods

2.1. Oligonucleotides and template DNA for the CDR amplification

The DP-47 germline gene was selected as the master framework and its sequence was camelised, by mutating three residues, according to Davies and Riechmann (1995). To build a V_H domain by the synthetic approach, oligonucleotides were synthesised and purified, as described previously (Söderlind et al., 1995). To shuffle the CDR into the master framework, oligonucleotides based on the DP-47 germline gene were used in a PCR. For each CDR amplification, an oligonucleotide pair was designed to amplify the CDR as well as to allow for one strand of the PCR product to be used in gene assembly (Fig. 1). A human cDNA library derived from peripheral blood lymphocytes was used as template for the CDR. The sequences of the oligonucleotides used in this procedure are shown in Table 1. These primers were designed to amplify the CDR, as defined by Kabat et al. (1991). In addition, the last codon of FR1 (V_H codon 30) was included in the amplification of CDR1, and the first codon of FR3 (V_H codon 66) was included in the amplification of CDR2.

2.2. PCR amplification of the CDR

A cDNA library constructed from peripheral blood B cells, producing IgM antibodies, was used as template (Ohlin et al., 1996). This library was PCR amplified, using primers amplifying intact genes belonging to the V_H1, 3, 4 and 5 families, prior to CDR amplification. All PCR were performed with reagents and AmpliTaq polymerase from Perkin Elmer (Foster City, CA) and thermocyclers from Cetus Corporation (Emeryville, CA). Each CDR was amplified in a 100 µl reaction, containing 200 µM of each dNTP, 1 µM of each primer (one of which is biotinylated), 2.5 U AmpliTaq, 0.1–1 ng V_H encoding DNA. The reaction profile used was: 94°C for 1 min, 55°C for 1 min, 72°C for 2 min for 30 cycles. To remove traces of the original template, the PCR product was purified from a 2% SeaPlaque low melting agarose gel (FMC Bioproducts, Rockland, ME), using QIAEX II gel extraction kit (QIAGEN, GmbH, Hilden, Germany).

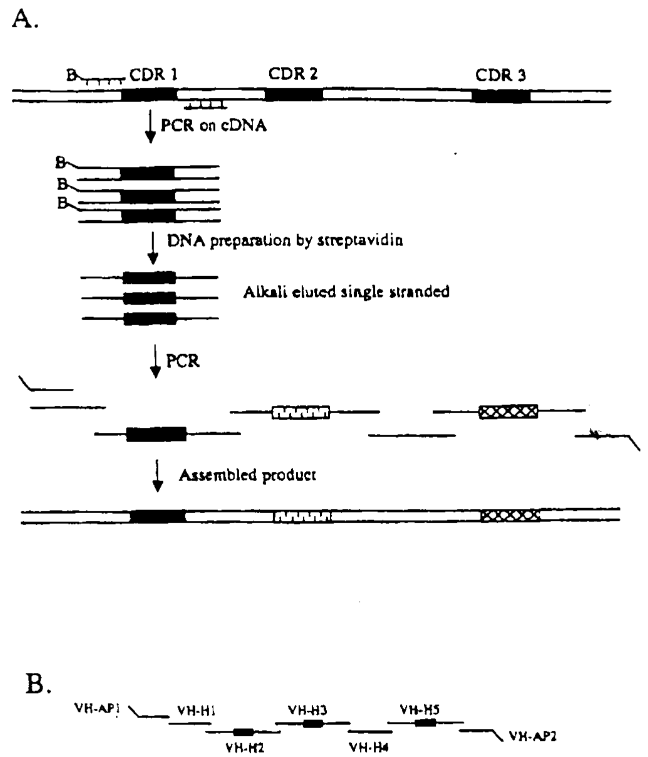

Fig. 1. The principle of CDR shuffling. CDR originating from different antibody genes can be randomly assembled into a given framework region. (A) Each CDR is amplified in a separate PCR reaction from, e.g. a cDNA library. One of the primers in the PCR amplification of each CDR is biotinylated. The PCR product is captured on a streptavidin-coated affinity column and single-stranded DNA encoding the CDR is prepared by alkali denaturation of the DNA and elution of the non-biotinylated strand. This single-stranded DNA is subsequently used in an assembly reaction, where CDR originating from different antibody genes will be randomly combined into a given framework region. (B) The location of the primers shown in Table 1. The primers V_H-AP1 and V_H-AP2 are amplification primers and the primers V_H-H1 and V_H-H4 are fully synthetic internal primers. The internal primers V_H-H2, -H3 and -H5 are prepared by amplification of naturally occurring CDR, as described above. The amplification primers are used to increase the copy number of each assembled gene, to introduce an N-terminal FLAG sequence and to provide restriction sites necessary for cloning. The boxes represent the CDR.

2.3. Preparation of single-stranded CDR-encoding DNA

To be able to utilise the PCR amplified CDR in an assembly reaction using overlapping oligonucleotides, single-stranded DNA was prepared initially. This was performed by affinity chromatography on the biotinylated strand, using an affini-tip (Genosys Biotechnologies, Inc, Pampisford, UK), as outlined in

...ces of primers used for the amplification of each CDR, the internal primers used in the assembly reaction and the outermost ...n primers used in the final amplification reaction

...r the amplification of the CDR
...r V_H-H2: 5′-biotin-GTC CCT GAG ACT CTC CTG TGC AGC CTC TGG ATT CAC CTT T
...er V_H-H2: 5′-TCC CTG GAG CCT GGC GGA CCC A
...ner V_H-H3: 5′-CGC CAG GCT CCA GGG AAG GAG AGG GAG GGG GTC TCA
...ner V_H-H3: 5′-biotin-GGA ATT GTC TCT GGA GAT GGT GAA
...mer V_H-H5: 5′-GAG CCG AGG ACA CGG CCG TGT ATT ACT GTG CAA GA
...rimer V_H-H5: 5′-biotin-GCG CTG CTC ACG GTG ACC AGG GTA CCT TGG CCC CA
...lly synthetic internal primers used in the assembly
...H1: 5′-GAG GTG CAG CTG TTG GAG TCT GGG GGA GGC TTG GTA CAG CCT GGG GGG TCC CTG AGA CTC TCC TGT
...H4: 5′-GGC CGT GTC CTC GGC TCT CAG GCT GTT CAT TTG CAG ATA CAG CGT GTT CTT GGA ATT GTC TCT GGA GAT GGT
...plification primers used in the assembly PCR
V_H-AP1: 5′-ACT CGC GGC CCA ACC GGC CAT GGC CGA GGT GCA GCT GTT GGAG
V_H-AP2: 5′-CAA CTT TCT TGT CGA CTT TAT CAT CAT CAT CTT TAT AAT CGC TGC TCA CGG TGA CCA

The location of each primer in the assembly reaction is shown in Fig. 1B.

Fig. 1. Briefly, the biotinylated PCR product was affinity captured on the streptavidin-coated matrix, and the column was washed to remove any remaining template DNA. The non-biotinylated strand was eluted by denaturing the DNA with alkali. The purification was performed essentially according to the manufacturer's instructions, except that all the washing steps using 1 × binding buffer were performed twice as many times as recommended. The eluted single-stranded DNA was subsequently used in the assembly reaction.

2.4. *Assembly of overlapping oligonucleotides into full-length V_H-genes*

Five overlapping internal oligonucleotides and two amplification primers were used in the assembly PCR (Söderlind et al., 1995). The three internal oligonucleotides encoding the CDR were prepared as described above, whereas all other oligonucleotides were synthesised and purified as described previously (Söderlind et al., 1995). The amplification primers were included to increase the copy number of each assembled gene and to provide relevant restriction sites. The genes were assembled in 25 μl reaction with 200 μM of each dNTP, 3 nM of each internal primer, 0.3 μM of each amplification primer, 0.625 U AmpliTaq. The reaction profile used was: 94°C for 1 min, 55°C for 1 min, 72°C for 2 min for 30 cycles. The assembled V_H genes were purified from a 1.5% SeaPlaque low-melting agarose gel using QIAEX II (QIAGEN) and re-amplified in 50 μl by PCR with primers that were identical to approx. 20 of the outermost bases of the PCR product. The reaction profile used was: 94°C for 1 min, 45°C for 1 min, 72°C for 2 min for 2 cycles, followed by 10 cycles with the profile: 94°C for 1 min, 55°C for 1 min, 72°C for 2 min. The re-amplified PCR product was gel-purified, as described above.

2.5. *Cloning of the assembled genes*

The assembled, re-amplified and purified genes were digested with *Nco*I and *Sal*I and cloned into the phagemid vector pEXmide5 for sequencing. This vector was originally constructed by introducing a cloning linker (Johansson and Söderlind, unpublished data) in the phage display vector pEXmide4 (Kobayashi et al., 1997). The genetic library was electroporated into *E. coli* XL-1 Blue and single colonies were used for plasmid preparation, using QIAGEN mini plasmid kit (QIAGEN).

2.6. *Sequencing and genetic analysis*

The complexity of the library was determined by sequencing both strands of 19 clones with Ready Reaction dye-terminator kit (Perkin Elmer), using specific primers. The sequences were also determined with Ready Reaction dye-primer kit (Perkin Elmer), using the M13 reverse primer. Sequence No. 5 was also sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer). The sequencing reactions were analysed using an ABI 377 automatic DNA sequencer and the sequences were evaluated using the Wisconsin sequence analysis package (version 8; Genetics Computer Group Inc., Madison, WI), using the FASTA algorithm. The origin of each CDR was determined by comparison with known variable gene sequences available in the March 1997 release of the V BASE sequence directory (Cook and Tomlinson, 1995).

3. Results and discussion

3.1. Selection of a master framework

The DP-47 germline gene was used as a master framework, together with in vivo formed CDR originating from different germline genes, for the construction of a gene library encoding soluble V_H domains. The master framework was selected since it folds easily and is well expressed in *E. coli* (Kobayashi et al., 1997), an important feature in antibody engineering (Plückthun and Pack, 1997). Primers suitable for this system have previously been defined and characterised with respect to hetero- and homodimer formation (Kobayashi et al., 1997). These data were applied in this study. Antibody fragments originating from the DP-47 germline gene are frequently selected in phage display systems (Griffiths et al., 1994). In addition, it is a germline gene that is often used in human immune repertoires (Huang et al., 1996; Ohlin and Borrebaeck, 1996), making it a good choice if selected clones are to be tolerated in in vivo applications.

3.2. The complexity of the library

A library of 9×10^6 members was created by transforming electrocompetent XL-1 Blue *E. coli*. To evaluate the complexity of the library, 19 randomly picked clones were sequenced. These sequences revealed that the CDR shuffled into the DP-47 framework originated from a diverse set of V_H genes. Also, the CDR1 and CDR2 found in each assembled gene frequently originated from different V_H genes, demonstrating that they had been combined randomly in each product (Table 2). Even though the V-region template was determined by sequencing to contain genes encoding members of V_H1, V_H3 and V_H4 families at a ratio of 3:12:4 ($n=19$), the sequences revealed that all CDR1 and CDR2 found in the CDR-shuffled V_H-library belonged to the V_H3 family. This is the family to which the DP-47 master framework belongs, too. The fact that only CDR originating from V_H genes belonging to the V_H3 family was found is, however, a consequence of our specific primer design (data not shown) and other V_H families have been amplified from the same DNA template using a different

Table 2
Demonstration of the complexity of the assembled, randomly selected genes[a]

	CDR1			CDR2			CDR3	
Sequence no.	Germline	No. of mutations	Homology to DP-47 aa	Germline	No. of mutations	Homology to DP-47 aa	No. of aa	Gene assembled in frame
DP-47			SSYAMS			AISGSGGSTYYADSVKGR		
2	DP-35	1/18	.D.Y..	DP-42	0/51	V.Y-..............	12	Yes
3	DP-49, DP-50	1/18	...G.H	DP-53	0/54	R.NSD.S..S........	13	No
5	DP-47	0/18	......	DP-51	7/54	Y.FRISSTV...E.....	11	Yes
6	DP-32	0/18	DD.G..	DP-47	0/54	....	10	No
7	DP-41	1/18	.N.G.H	DP-47	0/54	....	8	No
8	DP-32	0/18	DD.G..	DP-77	0/54	S..S.SSYI.........	12	Yes
9	DP-31	0/18	DD...H	DP-47	3/54	S.................	12	Yes
10	DP-31	0/18	DD...H	DP-31	3/54	G..WNSV.IV..E.....	14	No
12	DP-47	0/18	......	DP-60	2/54	G.....STI.........	18	No
13	DP-49, DP-50	0/18	...G.H	DP-35	4/54	Y..S..NTIN.....E..	16	No
14	DP-49, DP-50	0/18	...G.H	DP-35	2/54	Y..S..STI.........	7	No
15	DP-48	0/18	...D.H	DP-48	0/51	..G-TA.D...PG.....	13 2/3	No
16	DP-51, DP-77	0/18	...S.N	DP-47	1/54	V.................	13	No
17	DP-35	4/18	.D.SID	DP-31	0/54	G..WNS..IG........	10	Yes
19	DP-46, DP-61	2/18	IT...H	DP-53	7/54	R.NED.SD.N...A....	7	No
20	DP-31	0/18	DD...H	DP-47	7/54	S..SN.R......P....	7	Yes
21	DP-31	0/18	DD...H	DP-77	1/54	Y..S.SSYI.........	25	Yes
22	DP-41	0/18	.N.G.H	DP-35	0/54	Y..S..STI.........	15 1/3	No
23	DP-35	0/18	.D.Y..	DP-53	4/54	R.NSD.ST.G..E.....	10	Yes

[a]The origin of each CDR1 and CDR2, the number of nucleotide differences from the germline sequence and the length of CDR3 were determined. Each CDR1 in this evaluation includes codon 30, and each CDR2 includes codon 66, according to Kabat et al. (1991), as these sequences are not encoded by the synthetic primers. Included is also an amino acid (aa) comparison to CDR1 and CDR2 of DP-47 including the FR residues encoded by the template. The majority of the frameshift mutations resides in the synthetic primer region, and only in two cases in the CDR (sequences 15 and 22). Whenever a CDR showed maximal homology to several germline genes, functional genes described by Cook and Tomlinson (1995) were always chosen. Sequences no. 12, and 7 and 22 showed maximum homology to pseudogenes (DP-60 and DP-41, respectively).

primer design (Ohlin et al., unpublished data). Furthermore, the amplified CDR exhibited diversity and were, in all cases except three (one CDR each in sequence No. 7, 12 and 22), encoded by germline genes which are considered to be functional (Cook and Tomlinson, 1995). The fact that we can also amplify CDR from pseudogenes allows us to untap an even greater variability, since these particular CDR encode an in-frame product. Importantly, the shuffled CDR3 sequences were highly diverse in length (Table 2) and sequence (data not shown). In general, the length distribution (mean length = 12.3 aa) was close to lengths observed in antibodies developed in vivo (Ohlin and Borrebaeck, 1996; Wu et al., 1993). Mutations were found in the CDR despite the fact that the template used for CDR amplification originates from peripheral blood lymphocytes producing IgM antibodies. These observed mutations were not typical *Taq* polymerase-induced errors (Tindall and Kunkel, 1988), and their presence is consistent with recent findings of somatically mutated Ig-M encoding genes (Klein et al., 1997; Brezinschek et al., 1997).

3.3. *Analysis of the assembled V_H genes*

When further analysing the assembled genes, it was evident that 8/19 genes contained no frameshift mutations and encoded a complete product. When analysing the 11 non-functional genes, the deletion/insertions frequently occurred in sequences encoded by the synthetic oligonucleotides, i.e. primers not containing the CDR. In only two cases did the frameshift mutation occur in a CDR (sequences 15 and 22) which had frameshift mutations in its CDR3, suggesting that they originated from genes which had been improperly rearranged in vivo. The frameshift mutations residing in the framework regions were probably introduced during the synthesis of the synthetic oligonucleotides. In the synthesis of longer segments, the frequency of full-length, correctly deprotected product can drop dramatically. Despite the fact that the oligonucleotides used in the assembly had been purified on Oligonucleotide Purification Cartridge (Perkin Elmer), some genes still contained these deletions/insertions. Refinements in the technology of synthesising the oligonucleotides will improve the quality of the gene library and, thus, the power of this approach even further. Single-stranded DNA prepared from PCR products seems to have a much lower frequency of insertions/deletions, and the use of this type of single-stranded DNA will facilitate the assembly of functional genes.

3.4. *Conclusions*

In conclusion, we present data for a molecular design where gene segments formed in vivo can be shuffled into preselected framework regions, thus allowing the generation of an enormous diversity. This approach does not require engineering of restriction sites for the CDR shuffling and the variation introduced is the natural variation occurring in antibody genes. Our data support the proposal that this synthetic approach permits randomisation of all three CDR in one step. By using in vivo formed CDR, the length, sequence and combination of CDR can be varied at the same time. Thus, it does not require the design and synthesis of mutated synthetic oligonucleotides (Crameri and Stemmer, 1995; Virnekas et al., 1994), and by varying the CDR primers and the template source, it will also be possible to modify the complexity of the library. This approach will not only be applicable in antibody engineering, but also in the engineering of other proteins, where functional segments can be shuffled to obtain a protein with improved qualities.

Acknowledgement

This investigation was supported by grants from The National Research Council for Engineering Sciences, the National Board for Laboratory Animals and European Commission (BIO4-CT95-0252). We would like to thank Heléne Turesson for the synthesis of oligonucleotides.

References

Barbas III, C.F., Bain, J.D., Hoekstra, D.M., Lerner, R.A., 1992. Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. Proc. Natl. Acad. Sci. USA 89, 4457–4461.

Brezinschek, H.P., Foster, S.J., Brezinschek, R.I., Dörner, T., Domiati-Saad, R., Lipsky, P.E., Analysis of the VH gene repertoire. Differential effects of selection and somatic hypermutation on human peripheral $CD5^+/IgM^+$ and $CD5^-/IgM^+$ B cells. 1997. J. Clin. Invest. 99, 2488–2501.

Cook, G.P., Tomlinson, I.M., 1995. The human immunoglobulin VH repertoire. Immunol. Today 16, 237–242.

Crameri, A., Stemmer, W.P., 1995. Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wild-type sequences. Biotechniques 18, 194–196.

Davies, J., Riechmann, L., 1995. Antibody VH domains as small recognition units. Bio/Technology 13, 475–479.

Griffiths, A.D., Williams, S.C., Hartley, O., Tomlinson, I.M., Waterhouse, P., Crosby, W.L., Kontermann, R.E., Jones, P.T., Low, N.M., Allison, T.J., Hoogenboom, H.R., Nissim, A., Cox, J.P.L., Harrison, J.L., Zaccolo, M., Gherardi, E., Winter, G., 1994. Isolation of high affinity human antibodies directly from large synthetic repertoires. EMBO J. 13, 3245–3260.

Hayashi, N., Welschof, M., Zewe, M., Braunagel, M., Dübel, S., Breitling, F., Little, M., 1994. Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. Biotechniques 17, 310–315.

Hayden, M.S., Gilliland, L.K., Ledbetter, J.A., 1997. Antibody engineering. Curr. Opin. Immunol. 9, 201–212.

Hoogenboom, H.R., Winter, G., 1992. By-passing immunisation.

Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. J. Mol. Biol. 20, 381–388.

Huang, S.C., Jiang, R., Glas, A.M., Milner, E.C.B., 1996. Non-stochastic utilization of Ig V-region genes in unselected human peripheral B cells. Mol. Immunol. 33, 553–560.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., Foeller, C., 1991. Sequences of Immunological Interest, 5th ed. NIH Publication No. 91-3242. National Institutes of Public Health, Bethesda, MD.

Klein, U., Küppers, R., Rajewsky, K., 1997. Evidence for a large compartment of IgM-expressing memory B cells in humans. Blood 89, 1288–1293.

Kobayashi, N., Söderlind, E., Borrebaeck, C.A.K., 1997. Analysis of assembly of synthetic antibody fragments: expression of functional scFv with predefined specificity. Biotechniques 23, 500–503.

Ohlin, M., Borrebaeck, C.A.K., 1996. Characteristics of human antibody repertoires following active immune responses in vivo. Mol. Immunol. 33, 583–592.

Ohlin, M., Owman, H., Mach, M., Borrebaeck, C.A.K., 1996. Light chain shuffling of a high affinity antibody results in a drift in epitope recognition. Mol. Immunol. 33, 47–56.

Plückthun, A., Pack, P., 1997. New protein engineering approaches to multivalent and bispecific antibody fragments. Immunotechnology 3, 83–105.

Stemmer, W.P., 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389–391.

Söderlind, E., Vergeles, M., Borrebaeck, C.A.K., 1995. Domain libraries: synthetic diversity for de novo design of antibody V-regions. Gene 160, 269–272.

Tindall, K.R., Kunkel, T.A., 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 27, 6008–6013.

Virnekas, B., Ge, L., Plückthun, A., Schneider, K.C., Wellnhofer, G., Moroney, S.E., 1994. Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. Nucleic Acids Res. 22, 5600–5607.

Wu, T.T., Johnsson, G., Kabat, E.A., 1993. Length distribution of CDRH3 in antibodies. Proteins 16, 1–7.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., Arnold, F.H., 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. 16, 258–261.